DOI：10.13998/j.cnki.issn1002-1248.23-0691

中国农村留守儿童教育研究二十年
——基于结构主题模型

王　兴[1]，李叶叶[2*]，周天宇[3]，刘　峰[4]
（1. 中国科学院成都文献情报中心，成都　610299；2. 海南大学 管理学院，海口　570100；
3. 合肥工业大学 人工智能学院，宣城　242000；4. 濉溪县韩村中心学校小湖小学，淮北　235100）

摘　要：[目的 / 意义]随着国家扶贫政策和乡村振兴战略的提出，留守儿童教育问题引起了学者们的特别关注。然而，留守儿童教育的研究文献是零星和分散的，需要系统性的探讨。该研究的目的在于揭示留守儿童教育研究成果的主题内容及其演化规律，以为政策制定和实践提供更深入的理论支持。[方法 / 过程]该研究以 2002 年至 2023 年期间发表的 2 037 篇期刊文献摘要文本构建语料库，使用了基于结构主题模型方法对语料库进行主题建模，从主题多样性和动态性的角度来研究相关的现存文献及发展趋势。[结果 / 结论]通过主题建模的分析，最终确定了 8 个关键研究主题，分别是心理健康、留守儿童产生的前因、应对策略（宏观）、监护类型、综述类研究、家庭教育、媒介素养、应对策略（微观）。最后，在此基础上提出了未来的研究方向。

关键词：留守儿童；主题建模；结构主题模型；信息素养
中图分类号：G35　　文献标识码：A　　文章编号：1002-1248（2023）09-0043-14

引用本文：王兴，李叶叶，周天宇，等. 中国农村留守儿童教育研究二十年——基于结构主题模型[J]. 农业图书情报学报，2023，35(9)：43-56.

1 引　言

随着中国社会经济的快速发展，工业化和城市化进程的不断推进，城市创造了更多的就业机会，自 20 世纪 80 年代以来，农村劳动力不断向城市转移，农民工群体成为中国经济腾飞不可或缺的中间力量。与此同时，受户籍制度、经济水平的限制，农民工在城市中的住房、子女教育、子女照料等问题较难解决，很多农民工将子女留在农村，“农村留守儿童”群体逐渐形成[1]。教育部统计数据显示，2021 年中国小学、初中阶段农村留守儿童共计 1 198.65 万人。农村留守儿童缺失家庭教育、生活条件艰苦导致了诸多教育问题，如受教育难、学习成绩差、辍学率高等[2]。中国一直关注留守儿童问题，出台了《关于加强农村留守儿童关爱保护工作的意见》，并在十四五规划提出：“完

收稿日期：2023-08-06
基金项目：2022 年度中国科学院成都文献情报中心创新基金青年项目“基于 STM 主题模型的学科交叉主题识别与演化趋势研究”（E3Z0000303）
作者简介：王兴（1994-），男，博士，中级，研究方向为文本挖掘。周天宇（2002-），男，本科生，研究方向为人工智能、自然语言处理研究。刘峰（1976-），男，专科，研究方向为留守儿童教育研究
*通信作者：李叶叶（1994-），女，博士研究生，研究方向为教育管理研究。E-mail：liyeye0320@126.com

DOI：10.13998/j.cnki.issn1002-1248.23-0691

善留守儿童关爱体系，巩固义务教育控辍保学成果”[3]。教育不仅是留守儿童自身成长和人力资本的积累手段，也是巩固农村脱贫攻坚，阻断贫困传递的重要措施，更是实现乡村振兴的未来力量。

留守儿童的教育问题也引起了学术界的广泛关注，对现有文献进行更深入的组织、整合、综合和评估将为未来的研究和更具凝聚力的研究议题奠定基础。目前关于文献综述研究主要有 3 类：①传统的定性综述方法[4]。这种综述研究纳入分析的文章数量有限，依赖作者对该领域的积累。且研究结果较为主观，缺乏可复制性。②元分析法[5]。这种方法仅对留守儿童的定量研究适用。③基于 Citespace 的可视化分析方法[6]。这种方法产生的研究结果更为客观，可以复制，但该类方法不能对摘要等长文本信息进行挖掘。人工智能和机器学习技术使得大规模处理文本并进行深入分析成为可能，杜燕萍[7]运用 LDA 主题模型对教师队伍建设改革政策文本进行分析，挖掘出“激励导向”“职前培训”“师德师风建设”三大主题，并在此基础上提出政策建议；董伟[8]运用机器学习技术对高等教育国际化领域研究领域的文献进行主题识别与演化特征分析；王宇荣[9]运用 LDA 主题模型挖掘了关于留守儿童的新闻报道和网民的评论，研究了留守儿童媒介形象的差异化建构。由此可见，文本挖掘和机器学习技术逐渐受到教育学领域的接纳和重视。

在此背景下，我们的研究旨在通过系统的文献综述，对与农村留守儿童相关的教育问题形成系统而全面的理解。我们识别并分析了发表在期刊上 2 037 篇相关文献的摘要文本，应用基于主题模型的文本挖掘方法来揭示这一庞大文献中隐藏的语义结构和主题，并结合中国的情境特点提出未来的发展方向。

2 主题建模

2.1 背景介绍

主题建模是一种文本挖掘技术，用于发现给定文本集合的“隐藏”主题结构，可以快速有效地揭示大量非结构化数据集合中的主要主题或趋势。文本格式的学术论文本质上是非结构化的，因此，越来越多的学者使用主题建模来研究不同研究领域的主要话题趋势。Latent Dirichlet allocation（LDA）是最常用的主题建模方法，HWANG[10]运用 LDA 主题模型对 1981 年 1 月至 2022 年 3 月期间发表的有关技术在数学教育中的应用的 2 433 篇文献进行分析，从所选文章中提取关键术语和主题，确定七个研究主题以及这 7 个主题呈现出不同的研究趋势（稳定、波动、增加和减少）；ODDEN[11]对 1916 年以来在 *Science Education* 上发表的文章进行主题建模，挖掘出 21 个不同的主题，分为 3 个主题组：科学内容主题、以教学为中心的主题和以学生为中心的主题。田园[12]对在线教学评价内容进行主题建模，构建了在线教学用户需求指标体系。

然而，LDA 主题模型有很大的局限性。一方面，LDA 无法表示文档中主题和词随协变量的变化；另一方面，LDA 假设主题相互独立，不允许主题之间存在相关性。为了克服这些限制，本研究采用 Roberts[13]开发的结构主题模型（Structural Topic Model，STM）。STM 主题模型是 LDA 和相关主题模型的扩展，通过直接估计协变量对主题流行度或主题内单词使用的影响来解决 LDA 的局限性。与 LDA 相比，STM 具有 3 个明显的区别。首先，允许主题相关；其次，每个文档都有其在主题上的分布作为协变量的函数；第三，在一个主题中，单词分布也可能因协变量的变化而发生变化。先验分布是通过包含额外的协变量值来“结构化”的。由于包含用于主题分布建模的文档级协变量，STM 主题模型优于其他主题建模方法。

STM 主题模型最近才被引入主题建模家族，主要应用于社会学领域。汪顺玉[14]基于结构主题模型分析了“一带一路”学术话语的主题、变迁与差异；赵蓓[15]对 Twitter 中发表的关于北京冬奥会的文本进行结构主题建模，分析了社交机器人叙事与立场偏向。目前，尚未检索到 STM 主题模型在教育领域的应用。

2.2 结构主题模型

STM 也是一种概率生成模型，它将生成的文档定

义为具有不同流行度的隐藏主题的混合体。生成模型解释了特定文档是通过使用多个主题的数据生成过程而产生的。对于一个语料库，假设 K 个潜在主题形成一个潜在语义空间，其中每个词都可以表示为 K- 维度向量。此外，每个主题由具有不同流行度的一元词或二元词混合组成，这意味着每个词都有属于特定主题的概率。在这种情况下，对于一篇文档，所有主题的总主题比例为 1，对于特定主题，所有词的总词概率也为 1。最后，可以使用根据称为流行度的相应条件概率分数排名的前 n 个词来总结这些潜在主题。

图 1 为结构主题模型（STM）的原理图，其中，D_i 代表语料库中的文档；T_k 代表语料库潜在的主题；模型参数 θ_d 假定 STM 中的随机变量是从对数正态分布中提取的随机变量；X 代表协变量，可以是时间、类别等；β_x 表示主题–词汇的分布；w_j 表示构成主题的词汇。结构主题模型 STM 的生成是一个复杂的过程，更多的技术细节请参考[16]。

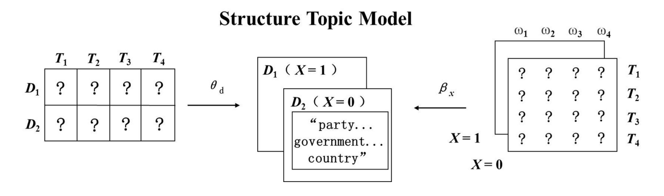


图 1　结构主题模型（STM）原理图

Fig.1 Structural topic model (STM)

3　数据与方法

3.1　研究框架

为了进行这项研究，我们采用结构主题模型（STM）对与农村留守儿童教育相关的文献进行建模，以形成连贯和全面的理解。图 2 为本研究的框架。

3.2　数据收集

本文使用中国知网（CNKI）收集已发表的关于农村留守儿童教育的期刊文献数据。首先，构造检索式：((TI=' 留守儿童 '+' 农村留守儿童 ') AND (TI=' 教育 ')) OR ((KY=' 留守儿童 ' +' 农村留守儿童 ') AND (KY=' 教育 '))，在标题或关键词中搜索“留守儿童 / 农村留守儿童”以及含有“教育”的文章。其次，对检索出的结果进行逐一审核，删除了与本研究无关的记录，如期刊征文。通过以上两步的处理，共检索出“农村留守儿童教育”相关的文献 2 037 篇，并将文献标题、摘要、关键词、作者、机构等信息存储在 Excel 表格中。时间跨度从 2003 年 1 月到 2023 年 1 月的 20 年。表 1 列举了发表论文数量超过 10 篇的 24 本期刊。

3.3　数据预处理

3.3.1　数据清洗

数据预处理的首先要去除文本中标点符号、数字和停用词（如“的”“因此”“所以”“本文”等无意义的词汇）。本研究通过 Python 语言编写程序，调用《哈工大停用词表》进行数据清洗。

3.3.2　专业词表构建

构建专业词表的目的是确保专业词汇不被分割，比如将“留守儿童”一词添加至专业词表，则“留守儿童”一词不会被程序切分为“留守 / 儿童”两个词。本研究根据所收集文献的关键词和留守儿童研究领域常用的概念构建专业词表，部分如表 2 所示。

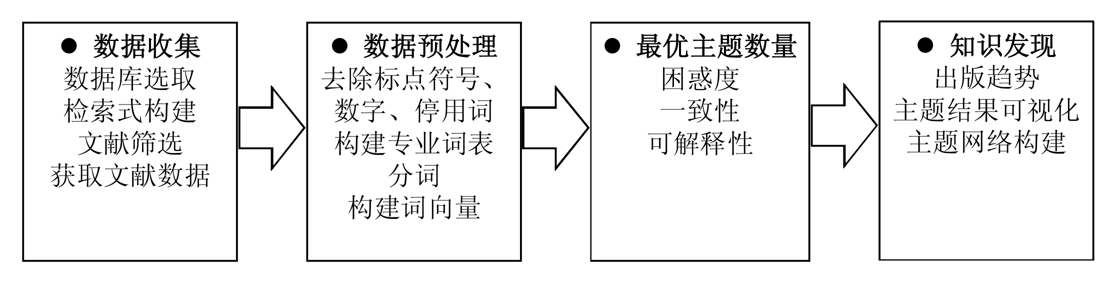


图 2　研究框架

Fig.2 Research framework

DOI：10.13998/j.cnki.issn1002-1248.23-0691

表 1 期刊分布

Table 1 Journal distribution

序号	期刊名称	论文数量/篇	序号	期刊名称	论文数量/篇
1	科学咨询（教育科研）	60	13	经济研究导刊	15
2	才智	49	14	科教文汇（中旬刊）	15
3	农村经济与科技	43	15	科学咨询（科技·管理）	15
4	教育教学论坛	37	16	西部素质教育	14
5	学理论	31	17	现代商贸工业	14
6	法制与社会	29	18	教育观察	13
7	亚太教育	29	19	文学教育（下）	13
8	现代教育科学	25	20	产业与科技论坛	12
9	现代交际	23	21	科教导刊（中旬刊）	12
10	科技信息	17	22	教育评论	11
11	教学与管理	16	23	教育现代化	11
12	中国教育学刊	16	24	新西部	11

表 2 专业词表（部分）

Table 2 Glossary of terms (partial)

序号	关键词	序号	关键词
1	留守儿童	21	城镇化
2	家庭教育	22	流动儿童
3	心理健康	23	健康教育
4	学校教育	24	寄宿制学校
5	隔代教育	25	健康成长
6	教育对策	26	生命教育
7	农民工	27	教育策略
8	监护人	28	性教育
9	义务教育	29	新农村建设
10	社会工作	30	留守学生
11	道德教育	31	乡村振兴
12	教育公平	32	思想教育
13	社会化	33	社会支持
14	精准扶贫	34	公共政策
15	学前教育	35	贫困地区
16	音乐教育	36	美术教育
17	城市化	37	心理教育
18	安全教育	38	教育困境
19	情感教育	39	教育缺失
20	临时监护人	40	社区教育

3.3.3 分　词

英文文本以空格作为单词的分割，但中文文本比较特殊，在建模前需要将文本进行分词处理，在词语之间加入空格，比如：“农村留守儿童教育问题”分词结果为“农村 留守儿童 教育问题”。本研究用 Python 编程语言，调用 Jieba 库完成分词处理。

3.3.4 词向量构建

词向量是计算机程序理解文本的语义的基础，需要将文本进行向量化表示。本研究采用 *TF–IDF*（术语频率 - 逆文档频率）用于量化每个术语在语料库中的相对重要性。

TF–IDF 值由两部分组成：词频（*TF*）和逆文档频率（*IDF*）。*TF* 评估一个术语的出现频率。第 i 个关键词的 *TF–IDF* 值可以通过公式（1）和公式（2）计算。本研究使用 R 语言“quanteda”包[17]计算 *TF–IDF* 值。

$$TF-IDF_i = TF_{ij} * IDF_i \tag{1}$$

其中，TF_{ij} 代表第 i 个关键词在第 j 个文档中的频率。

$$IDF_i = log\frac{N+1}{N_i+1} + 1 \tag{2}$$

3.4 最优主题数量

STM 最重要的步骤指定主题数 K。确定 K 的值是为了实现对模型结果的实质性解释，而不是为了模型的最大化拟合[18]。一致性和排他性是模型结果解释能力的两个关键指标。高一致性表明主题中的热门词在

文档中同时出现的可能性较高，高排他性意味着该主题的高概率词通常不会出现在其他主题中[19]。本研究运用 R 语言 STM 包对语料库进行主题建模，提取了主题数量 K 从 3 到 20 共计 18 个模型的一致性和排他性评分，具体结果绘制在图 3。

从图 3 中可以看出，当主题数量较大时，模型表现出较高的排他性和较低的语义一致性，集中在图 3 左上区域；当主题数量较小时，模型表现出较低的排他性和较高的语义一致性，集中在图 3 右下区域；当主题数量为 6~9 时，模型的语义一致性和排他性均比较高。最终，我们选择了主题数量为 8，因为这种设置可以产生语义上一致性和排他性以及可解释性最强的结果。

4 研究结果

4.1 描述性结果

图 4 描绘了年度发表的期刊文章和包含报纸、会议的文章总量，可以观察到自 2004 年以来留守儿童教

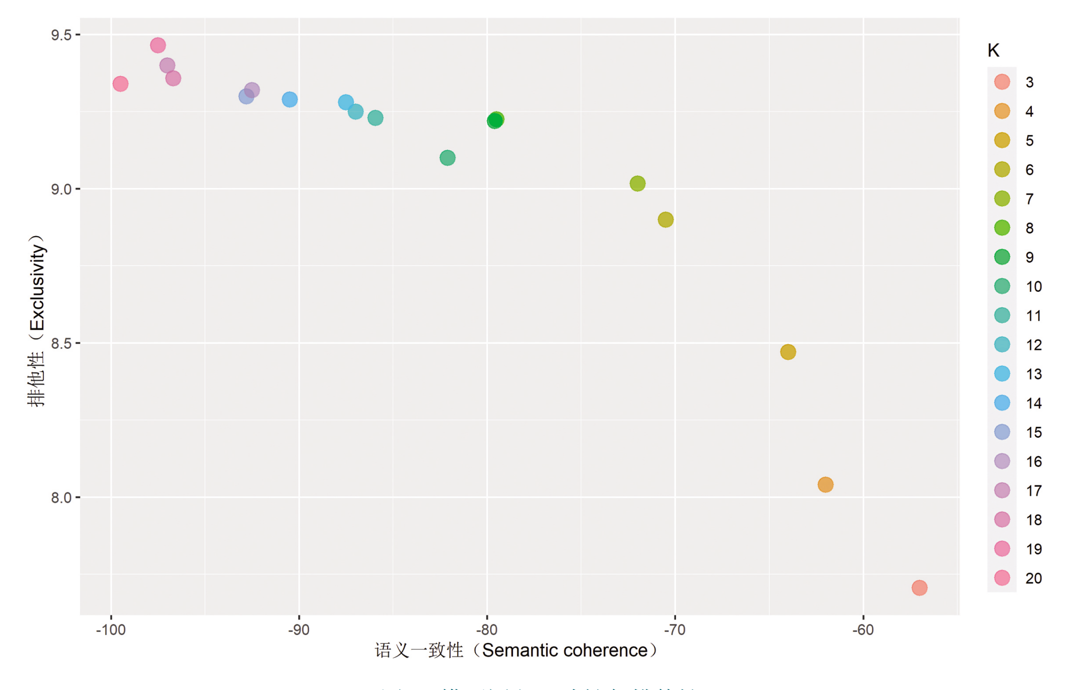


图 3 模型语义一致性与排他性

Fig.3 Model semantic consistency and exclusivity

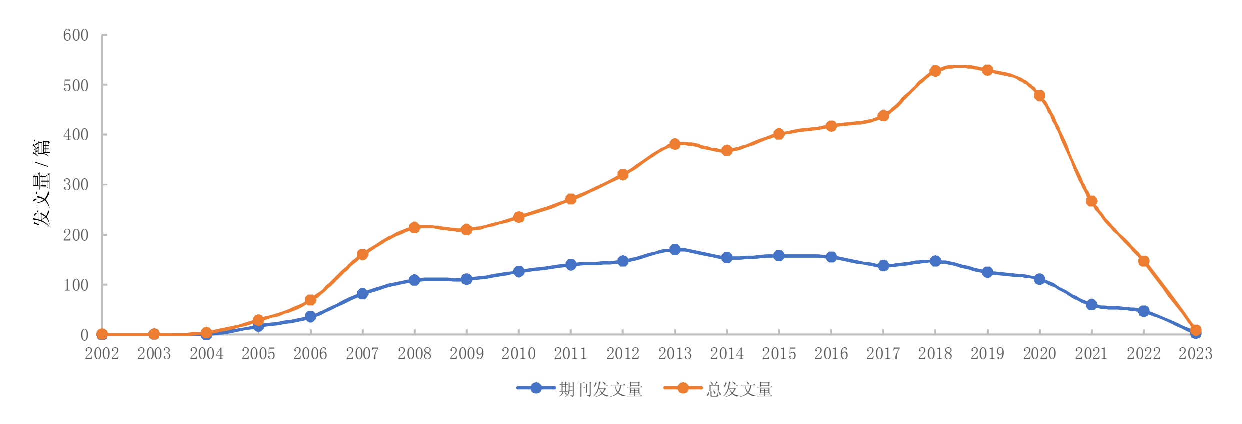


图 4 农村留守儿童教育研究年度分布

Fig.4 Annual distribution of papers published on education for rural left-behind children

DOI：10.13998/j.cnki.issn1002-1248.23-0691

育研究的快速增长。2003 年在期刊上发表首篇与留守儿童教育的文章 2013 年增长到 170 篇达到峰值，随后出现回落。2014 年之前，期刊文献和总文献数量趋势保持一致，2014 年之后，期刊文献数量逐渐减少，包含报纸、会议的总发文量继续增加，至 2019 年达到峰值 527 篇，随后迅速下降。

4.2 结构主题模型结果

STM 根据语义关联将文档语料库中的特定关键词分配到潜在主题，在这种情况下，每个主题都可以用小部分关键词来概括。为了确定每个主题的具体内容，我们提取了与每个主题相关的前 20 个关键词，具体结果在表 3 中呈现。STM 主题模型确定了与留守儿童教育问题相关的 8 个研究主题，进一步归纳为 7 个类别，分别是：心理健康、留守儿童产生的前因、应对策略（主题 3、主题 8）、监护类型、综述类研究、家庭教育、媒介素养。

4.2.1 心理健康

主题 1 涉及留守儿童的心理健康问题。留守儿童与父母分离时处于身心发展的重要阶段，心理健康问题是农村留守儿童教育中普遍存在的问题[20]。刘红艳[21]研究发现，父母外出务工 4 个月以上会对留守儿童的心理健康产生负面影响，父母返乡会缓解部分负面影响，但无法修复处于青春期的留守儿童的心理问题。王鑫强[22]以 5 个不同经济带的 1 066 名留守儿童为样本，将儿童心理健康按照积极 - 消极、内化 - 外化划分为 4 个维度，研究发现留守儿童在积极心理健康维度，如生活满意度、快乐感等维度均低于非留守儿童。留守儿童的心理健康也会受经济条件的影响，贫困地区的留守儿童心理健康水平较低。留守经历也会对心理健康产生长期效应，杨雪岭[23]研究发现，有留守经历的大学生心理韧性水平可能较低，存在较高的自杀风险。心理干预已被证实是提升留守儿童心理健康水平的有效手段，白勤[24]涉及并实施了“4+1”培养干预模式，显著提升留守儿童心理健康水平。

4.2.2 留守儿童现象产生的前因

主题 2 涉及到留守儿童现象产生的前因。农村劳动力转移、城乡二元体制是造成留守儿童现象的直接

表 3 STM 主题模型结果

Table 3 Classification results based on STM topic model

主题序号	研究主题	关键词
主题 1	心理健康	心理健康、农村、学生、心理、知识、干预、性教育、情感教育、留守儿童、青春期、生命教育、安全教育、非留守儿童、生命、积极、案件、青春期，对照组、活动、老师
主题 2	留守儿童产生的前因	务工、进城、期望、农民工、子女、限制、农民、城市、不得不、条件、市场经济、只能、涌向、离开、大批、子女教育、未成年、留在、隔代教育、家乡
主题 3	应对策略（微观）	帮扶、精准扶贫、机制、建立、体系、治理、寄宿制学校、乡村振兴、社区教育、创新、责任、扶贫、教育扶贫、服务体系、多方、教育公平、关爱教育、教育资源、政府、职能
主题 4	监护类型	一方、缺少、长期、双方、父母、外出、母亲、孩子、父亲、照顾、呵护、外出务工、未成年人、远离、分离、得不到、容易、快乐、亲戚、在外
主题 5	综述类研究	对策、问题、原因、引起、社会各界、关注、解决、阐述、广泛、分析、存在、现状、措施、探讨、社会转型、以期、普遍、已经、解决问题、成因、建议
主题 6	家庭教育	思想、道德、家庭教育、缺位、不良、弱化、观念、文化、人格、弥补、表现、性格、学校教育、缺失、习惯、社会化、思想道德、培养、意识、价值观
主题 7	媒介素养	媒介、大学生、社会工作、媒介、电视、访谈、文献、手机、介入、高校、媒介素养、调研、维度、支教、梳理、实证、工具、技术、媒体、信息化
主题 8	应对策略（宏观）	转移、学前教育、人口、劳动力、大规模、进程、教育权、新农村建设、现代化、城市化、城乡二元、经济社会、社会主义、加快、剩余劳动力、改革、和谐社会、未来、户籍制度、就业

原因[25]，不可忽视的是，经济压力大、子女教育难、自身流动性等是农民工选择将子女留在家乡的根本原因。在一项对留守儿童父母选择将子女留守在家原因的调查中发现，63.7%是因为“家庭经济压力重”[26]；学费贵、生活成本高、工作时间紧张、流动性大等也是农民工没有选择将孩子带在身边的主要原因[27]。

4.2.3 应对策略

主题 3 和主题 8 均体现了留守儿童教育问题的应对策略。为了更深入的挖掘两者的差异，我们计算两个主题关键词的概率差异并进行可视化，图 5 体现了二者的词汇差异。可以发现，主题 3 主要关注如何解决当前留守儿童教育问题，关注的角度更加微观；主题 8 注的角度更加宏观，关注农村未来的建设和发展。任运昌[28]基于 13 年的跟踪研究，提出留守儿童教育问题的应对策略和缓解策略，面对当前留守儿童的教育问题，应该落实政府行动主体的责任，开展“社区共育”“结对帮扶”“还原家教”“家校一体”的工作模式；与此同时，系统改善农村文化生态环境，防止留守儿童“污名化”，落实法律保护与法制教育，加强农村教师队伍建设，构建以寄宿制学校为中心的教育监护体系才是解决农村教育问题重中之重。在社会各界关注和应对留守儿童问题的过程中，也有学者敏锐的发现留守儿童的“污名化”，认为留守儿童和非留守儿童并无区别，而是成长环境造成的差异，是当前的社会结构产生了大量的留守儿童，需要从源头解决产生留守儿童的社会结构问题、呼吁实现城乡协同、乡村振兴、权利平等、让留守儿童的父母在家乡就能安居乐业，让农村留守儿童不再“留守”[29]。



图 5 主题 3 与主题 8 词汇差异

Fig.5 Vocabulary differences between topic 3 and topic 8

4.2.4 留守儿童监护类型

主题 4 为留守儿童监护类型。叶敬忠[30]对中西部 10 个农村社区的留守儿童的监护情况进行调查，将留守儿童的监护类型分为单亲监护（父亲 / 母亲一方监护）、隔代监护（爷爷、奶奶、外公、外婆等祖辈监护）、上代监护（亲戚、老师等成年人监护）及同辈监护（年龄稍大的哥哥、姐姐监护，或独居）4 种监护类型。范兴华[31]对 593 名留守儿童和 353 名父母共同监护的儿童进行研究，发现隔代监护、上代监护的留守儿童的违纪、多重行为、行为适应困难等问题行为维度高于父母监护的一般儿童。留守儿童的不同监护类型也会对留守儿童的心理健康产生不同的影响[32]。

4.2.5 综述类研究

主题 5 主要体现了综述类的研究。该类研究对现有的文献进行梳理，分析留守儿童教育问题产生的前因后果以及应对策略。我们已经在引言部分将现有的关于留守儿童教育问题的综述研究分为 3 类：传统的定性综述方法、元分析法和基于 Citespace 的可视化分析方法，在此不再赘述。

4.2.6 家庭教育

主题 6 主要讨论留守儿童家庭教育缺失的问题。留守儿童面临的最大的问题是家庭教育的缺失，家庭教育与儿童的人格、品格、性格等息息相关，留守儿童家庭教育的缺失会引起学习成绩、思想道德、心理素质、身体发育等方面出现问题[33]。家庭是儿童最重要的社会化场所[34]，家庭教育的缺失会加大同辈群体对留守儿童价值观和社会化的影响，一方面，价值观的形成受遗传因素和环境因素的共同作用，由于缺乏父母的影响，留守儿童会将同龄群体的行为方式作为评价自己和他人的标准，作为自己的社会观和价值观的依据[35]；另一方面，缺乏父母的正确引导，留守儿童容易加入不良群体、在社会化过程中偏离正确的价值观，从而导致不良行为甚至违法犯罪的发生[36]。家庭教育在儿童成长过程中发挥着非常重要的功能，面对留守儿童面临家庭教育缺失的现实问题，相关研究呼吁学校教育帮助留守儿童构建精神世界，树立正确

DOI：10.13998/j.cnki.issn1002-1248.23-0691

的道德管和价值观。随着农村规模学校和寄宿制学校的增多，学校和老师成为留守儿童成长的主要环境和陪伴者，将会在留守儿童的精神世界发挥更大的作用[37]。

4.2.7 媒介素养

媒介素养是指对媒介内容所形成的认知、处理和参与能力，媒介是儿童社会化发展的一个重要影响因素[38]，电视、电脑、手机等是留守儿童获取学习和生活的必要信息，与父母保持交流的重要窗口，对留守儿童的认知、学习、道德品质、人际关系等均有重要影响[39]。数字化环境下，农村留守儿童的媒介素养备受关注，研究发现农村留守儿童普遍缺乏媒介素养知识，家庭媒介环境较为单一，学校有关媒介素养的教育和设施比较缺乏[40]，媒介素养已被证实与留守儿童的学习能力正相关[41]。随着计算机技术的发展，智能手机已经走进部分留守儿童的生活，视频、语音、图像等沟通方式的普及使得留守儿童与父母形成了便捷稳定的情感沟通渠道，可以弥补留守儿童社会化过程中家庭教育缺失的部分负面影响。王清华[42]研究发现，智能手机通过“数字代偿”机制在一定程度上弥补留守儿童家庭氛围、教育和互动等方面的不足，使在他乡的在虚拟情境下参与留守儿童情感社会化的过程，但并不能完全弥补父母的陪伴。

4.3 主题发展趋势

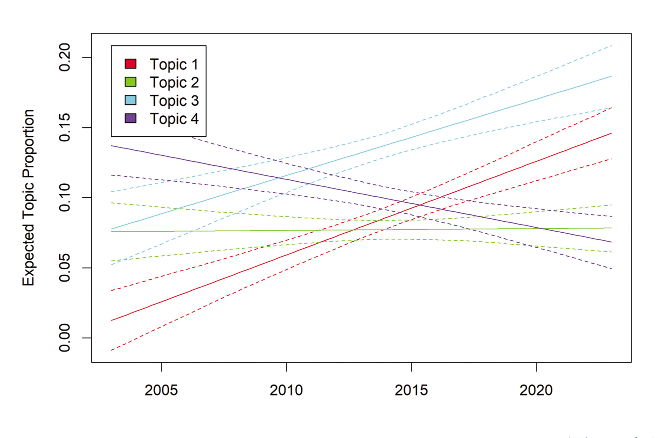
结构主题模型的主要优点之一是可以研究协变量和主题之间的相互作用。通过将文献的刊出年份设置为协变量，可以比较主题随时间的变化趋势。所有主题随时间动态如图 6 所示，可以看到每个话题的热度变化。过去 20 年中，主题 1（心理健康）、主题 3（应对策略：微观）、主题 7（媒介素养）呈上升趋势，主题 2（前因）、主题 6（家庭教育）热度基本不变，主题 4（监护类型）、主题 5（综述）、主题 8（应对策略：宏观）呈下降趋势。根据该图，可以识别热点和冷话题。热门话题具有增加趋势，冷门话题具有减少趋势。未来可以进一步分析这些变化趋势产生的原因。

5 结果与讨论

5.1 研究结果

留守儿童教育领域目前进行了大量的研究工作，学术文章数量众多，因此准确识别主要研究主题和趋势是一项耗时且具有挑战性的任务。本研究使用结构主题模型（STM）对留守儿童教育领域的文献进行综述，涵盖 777 种期刊的 2 037 篇论文，时间跨度从 2002 年 1 月到 2023 年 1 月。本研究呈现了留守儿童教育研究的整体概览和知识结构。基于主题模型的语义一致性和排他性确定了 8 个研究主题，并将文献的发表年份作为协变量确定主题的演化趋势。

结果表明，留守儿童教育领域的学术关注与留守儿童产生的前因、后果和作用机制有关，在过去的 20 年里，心理健康、应对策略、媒介素养 3 个主题的研



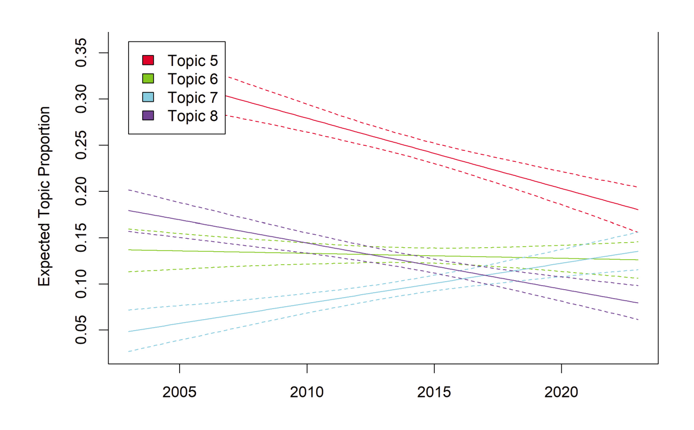


图 6 主题变化趋势

Fig.6 Trend of topic changes

究趋势不断上升。未来的研究应重点关注留守儿童家庭教育的缺失带来的问题，通过政府、学校、社区、家庭的共同努力，为留守儿童营造更良好的生存环境。

综合以上研究结果，可以看出留守儿童教育涉及的主题多样且动态变化。心理健康问题是留守儿童教育研究中的一大关注点，研究者们认识到留守儿童在心理健康方面面临的风险，并通过探讨心理健康问题的原因、影响因素和干预策略等，为留守儿童的心理健康提供了理论和实践支持。

此外，研究者们也关注了留守儿童产生的前因，包括经济、家庭、社会等多层面的因素，认识到留守儿童问题的根本原因较为复杂，需要综合多方面的因素进行研究和干预。

在应对策略方面，研究者们从宏观和微观两个层面进行了研究。宏观层面包括政府、社会组织和学校等的应对策略，如政策制定、社会服务体系建设和学校支持等。微观层面包括留守儿童自身的应对策略，如情绪调节、自我管理和社交能力等。这些研究为留守儿童教育提供了多层面、多维度的干预和支持策略，有助于促进留守儿童的全面发展和健康成长。

5.2　未来展望

基于对现有文献的综述和分析，未来的研究可以从以下几个方向展开。

深入研究留守儿童心理健康问题。进一步深入探究留守儿童心理健康问题的机制、风险因素和保护因素，包括心理健康评估、心理健康干预和心理健康政策的研究。同时，也可以考虑心理健康问题与其他因素如家庭背景、社会支持和文化因素的相互作用，深化对留守儿童心理健康问题的理解。

深入探究留守儿童产生的前因。进一步深入研究留守儿童产生的原因，包括经济、家庭、社会等多层面的因素，如留守儿童家庭背景、家庭教育方式、社会文化背景等。通过深入研究留守儿童产生的前因，可以为制定针对性的干预策略提供更为科学的依据。

深入研究留守儿童的应对策略。进一步深入研究留守儿童自身的应对策略，包括情绪调节、自我管理、社交能力等方面的研究。可以通过多种研究方法如心理测量、实证研究和长期追踪研究等，深入了解留守儿童应对困境和逆境的方式和效果，并为提供有针对性的干预策略提供科学依据。

强调家庭教育的重要性。留守儿童的家庭环境对其教育和发展具有重要影响。未来研究可以进一步关注留守儿童家庭教育的特点、问题和干预策略，包括家庭教育方式、家庭支持、亲子关系等方面的研究。同时，可以考虑将家庭教育纳入到留守儿童教育干预的综合模式中，强调家庭教育在留守儿童教育中的重要性。

关注媒介素养的发展。随着数字化媒体的普及，留守儿童在媒体环境中的成长和媒介素养的发展也日益受到关注。未来研究可以深入探究留守儿童在媒体使用方面的行为、习惯和影响，包括社交媒体使用、网络游戏行为、在线学习等方面的研究。同时，可以关注留守儿童媒介素养的培养策略和效果评估，为留守儿童在数字化媒体环境中的健康成长提供科学依据。

深入研究监护类型对留守儿童的影响。留守儿童的监护类型对其教育和发展产生重要影响。未来研究可以深入探究不同监护类型对留守儿童的影响，包括父母、祖父母、其他亲属或社会机构等监护者对留守儿童的支持、照顾和教育方式的研究。同时，可以探讨监护类型对留守儿童心理健康、学校适应、社会适应等方面的长期影响，并提出相应的干预策略。

加强综述类研究的方法和应用。综述类研究在留守儿童教育研究中具有重要地位。未来研究可以进一步加强综述类研究的方法和应用，包括系统性文献回顾、荟萃分析、元分析等方法的应用，从而更全面地总结和归纳留守儿童教育研究的现有成果，并为未来研究和实践提供更为可靠和有力的证据。

考虑微观层面的应对策略。除了宏观层面的应对策略外，留守儿童在日常生活中也会运用各种微观层面的应对策略。未来研究可以深入探究留守儿童在日常生活中的应对策略，包括个体层面的情绪调节、自我保护、社交支持等方面的研究。这将有助于更好地理解留守儿童在面对困境和压力时的实际应对行为，

DOI：10.13998/j.cnki.issn1002-1248.23-0691

为留守儿童提供更有效的心理健康支持和干预策略。

此外，未来研究也应关注留守儿童教育政策和实践的评估和改进，推动留守儿童教育政策的落地和实施，为留守儿童提供更好的教育和发展机会。同时，跨学科和跨领域的研究合作也是未来研究的趋势，如教育学、心理学、社会学、媒体学、社会工作等学科之间的合作，以促进留守儿童教育研究的深入和发展。

5.3 研究贡献与不足

本研究的贡献如下。本研究使用基于机器学习方法的 STM 主题模型对文献进行挖掘，可以克服基于 LDA 主题模型研究的局限性。人工智能的方法还允许学者从大量研究论文中快速获取全面的信息。本研究通过整合文档元数据信息，进一步系统地揭示了主题随时间的演化趋势，有助于确定随着时间的推移改变留守儿童教育研究重点的潜在趋势。

然而，该方法也有一定的局限性。首先，在本文中，我们从文献中仅收集了文章摘要而不是全文。因此，本研究中得到的这些主题可能不是完整研究文章中所含信息的完整表示。未来可以基于全文本挖掘技术对留守儿童教育领域进行更深入的挖掘。其次，在本文的检索中，我们重点考虑了“留守儿童”“农村留守儿童”以及“教育”的组合，这会损失一些文献，如以“留守子女”为关键词的研究，未来，可以通过构建机器学习分类模型，运用可解释性机器学习方法对论文分类中关键词的选择进行甄选[43]。最后，本研究并未将研究生学位论文纳入分析的范畴。研究生学位论文通常会修改成小论文进行发表，与期刊论文存在重叠，因此我们未将其纳入考虑范围。然而，也需注意到存在一些高质量的研究生学位论文。未来的研究可以考虑基于作者 - 论文相似度的计算，将未被修改为小论文发表的学位论文内容纳入统计分析中。

参考文献：

[1] 杨菊华, 段成荣. 农村地区流动儿童、留守儿童和其他儿童教育机会比较研究[J]. 人口研究, 2008, 32(1): 11-21.

YANG J H, DUAN C R. A comparative analysis of educational opportunities for migrant, stay and other children in rural China[J]. Population research, 2008, 32(1): 11-21.

[2] 王素芳, 谭清安. 乡村振兴背景下贫困地区农村面向青少年的数字文化环境建设——基于留守儿童网络使用和图书馆服务需求的调查分析[J]. 农业图书情报学报, 2022, 34(1): 16-37.

WANG S F, TAN Q A. How to create a new digital culture environment for teenagers in poor rural areas under the background of rural revitalization: Based on a study on internet use behavior and library services needs of left-behind children[J]. Journal of library and information science in agriculture, 2022, 34(1): 16-37.

[3] 中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要 _ 滚动新闻 _ 中国政府网[EB/OL].[2023-02-12]. http://www.gov.cn/xinwen/2021-03/13/content_5592681.htm.

[4] 戈静怡, 覃子怡, 杨杭, 等. 农村留守儿童家庭教育研究综述与展望——以隔代抚养家庭为例[J]. 教育观察, 2020, 9(31): 24-26.

GE J Y, QIN Z Y, YANG H, et al. Review and prospect of family education research on left-behind children in rural areas - Taking families raised by different generations as an example[J]. Survey of education, 2020, 9(31): 24-26.

[5] 白慧慧, 王雨晴, 孙婉靖. 留守儿童心理韧性的元分析[J]. 心理与行为研究, 2022, 20(2): 261-267.

BAI H H, WANG Y Q, SUN W J. The resilience status of left-behind children: A meta-analysis[J]. Studies of psychology and behavior, 2022, 20(2): 261-267.

[6] 李艺敏, 李永鑫. 留守儿童研究: 基于 CiteSpace 的可视化分析[J]. 西南民族大学学报(人文社会科学版), 2014, 35(8): 181-189.

LI Y M, LI Y X. Research on left-behind children: Visual analysis based on CiteSpace[J]. Journal of southwest university for nationalities (humanities and social science), 2014, 35(8): 181-189.

[7] 杜燕萍. 基于 LDA 主题建模的教师队伍建设改革政策文本分析[J]. 系统科学与数学, 2022, 42(6): 1411-1422.

DU Y P. Text analysis on the reform policy of teachers construction in the new ERA - Based on LDA modeling analysis[J]. Journal of systems science and mathematical sciences, 2022, 42(6): 1411-1422.

[8] 董伟, 董思遥, 张亚楠. 我国高等教育国际化领域研究主题识别与演化特征分析——基于文本挖掘的视角[J]. 教育理论与实践, 2021, 41(36): 3-7.

DONG W, DONG S Y, ZHANG Y N. An analysis of identification and evolution characteristics of research topics in the internationalization of higher education in China – A perspective based on text mining[J]. Theory and practice of education, 2021, 41(36): 3–7.

[9] 王宇荣, 项国雄. 留守儿童媒介形象的差异化建构——基于新闻、评论文本的分析[J]. 现代传播(中国传媒大学学报), 2019, 41(6): 17–23.

WANG Y R, XIANG G X. Differentiated construction of left–behind children's media image – Based on the analysis of news and commentary text[J]. Modern communication (journal of communication university of China), 2019, 41(6): 17–23.

[10] HWANG S, FLAVIN E, LEE J E. Exploring research trends of technology use in mathematics education: A scoping review using topic modeling[J]. Education and information technologies, 2023, 28(8): 10753–10780.

[11] ODDEN T O B, MARIN A, RUDOLPH J L. How has Science Education changed over the last 100 years? An analysis using natural language processing[J]. Science education, 2021, 105(4): 653–680.

[12] 田园, 宫婷婷. 基于LDA模型的在线教学需求数据主题挖掘研究[J]. 情报科学, 2021, 39(9): 110–116.

TIAN Y, GONG T T. Topic mining of online teaching demand data based on LDA model[J]. Information science, 2021, 39(9): 110–116.

[13] ROBERTS M E, STEWART B M, TINGLEY D. Stm: An R package for structural topic models[J]. Journal of statistical software, 2019, 91(2): 1–40.

[14] 汪顺玉, 陈瑞哲. 基于结构主题模型的"一带一路"倡议国际学术话语文本挖掘: 主题、变迁与差异[J]. 外语教学, 2022, 43(4): 8–14.

WANG S Y, CHEN R Z. A structural topic modeling of international academic discourse on the Belt and Road Initiative: Themes, trends and differences[J]. Foreign language education, 2022, 43(4): 8–14.

[15] 赵蓓, 张洪忠. 有关北京冬奥会的社交机器人叙事与立场偏向——基于Twitter数据的结构主题模型分析[J]. 新闻界, 2022(5): 62–70.

ZHAO B, ZHANG H Z. Social bots' narrative and position tendencies about the Beijing Winter Olympics –Structural topic model (STM) analysis based on twitter data[J]. Journalism and mass communication, 2022(5): 62–70.

[16] ROBERTS M E, STEWART B M, AIROLDI E M. A model of text for experimentation in the social sciences[J]. Journal of the American statistical association, 2016, 111(515): 988–1003.

[17] BENOIT K, WATANABE K, WANG H Y, et al. Quanteda: An R package for the quantitative analysis of textual data[J]. Journal of open source software, 2018, 3(30): ID 774.

[18] CHANG J, BOYD–GRABER J, GERRISH S, et al. Reading tea leaves: How humans interpret topic models[J]. Advances in neural information processing systems 22 – proceedings of the 2009 conference, 2009, 22: 288–296.

[19] ROBERTS M E, STEWART B M, TINGLEY D, et al. Structural topic models for open–ended survey responses[J]. American journal of political science, 2014, 58(4): 1064–1082.

[20] 赵峰. 农村留守儿童心理健康状况及教育对策[J]. 首都师范大学学报(社会科学版), 2010(3): 128–130.

ZHAO F. Mental health status of left–behind children in rural areas and educational countermeasures[J]. Journal of capital normal university (social sciences edition), 2010(3): 128–130.

[21] 刘红艳, 常芳, 岳爱, 等. 父母外出务工对农村留守儿童心理健康的影响: 基于面板数据的研究[J]. 北京大学教育评论, 2017, 15(2): 161–174, 192.

LIU H Y, CHANG F, YUE A, et al. The impact of parental migration on the mental health of rural left–behind children[J]. Peking university education review, 2017, 15(2): 161–174, 192.

[22] 王鑫强, 霍俊妤, 张大均, 等. 农村留守与非留守儿童的心理健康、虐待经历比较及其关系研究——基于两维四象心理健康结构的分析与对策建议[J]. 中国特殊教育, 2018(1): 58–64.

WANG X Q, HUO J Y, ZHANG D J, et al. The relationship between mental health and child abuse and their difference between left–at–home children and those living with their parents in rural areas: An analysis based on the two–dimensional four–fact mental health structure[J]. Chinese journal of special education, 2018(1): 58–64.

[23] 杨雪岭, 冯现刚, 崔梓天. 大学生的留守经历与心理韧性、心理病理症状[J]. 中国心理卫生杂志, 2014, 28(3): 227–233.

YANG X L, FENG X G, CUI Z T. Relation of left–behind experience to psychological resilience and psychopathological symptoms in

DOI：10.13998/j.cnki.issn1002-1248.23-0691

couege students[J]. Chinese mental health journal, 2014, 28(3): 227-233.

[24] 白勤, 林泽炎, 谭凯鸣. 中国农村留守儿童培养模式实验研究——基于现场干预后心理健康状况前后变化的数量分析[J]. 管理世界, 2012(2): 62-72.

BAI Q, LIN Z Y, TAN K M. An experimental study on the training mode of left-behind children in rural areas of China - Based on the quantitative analysis of the changes of mental health status after on-the-spot intervention[J]. Management world, 2012(2): 62-72.

[25] 段成荣, 吕利丹, 王宗萍. 城市化背景下农村留守儿童的家庭教育与学校教育[J]. 北京大学教育评论, 2014, 12(3): 13-29, 188.

DUAN C R, LV L D, WANG Z P. Research on left-behind children's home education and school education[J]. Peking university education review, 2014, 12(3): 13-29, 188.

[26] 王艳波, 吴新林. 农村"留守孩"现象个案调查报告[J]. 青年探索, 2003(4): 7-10.

WANG Y B, WU X L. A case investigation report on the phenomenon of "left-behind children" in rural areas[J]. Youth studies, 2003(4): 7-10.

[27] 段成荣, 杨舸, 王莹. 关于农村留守儿童的调查研究 [J]. 学海, 2005(6): 25-29.

DUAN C R, YANG G, WANG Y. Investigation and research on left-behind children in rural areas[J]. Journal of Xuehai, 2005(6): 25-29.

[28] 任运昌. 农村留守儿童教育问题的当前态势、应对模式与缓解策略——基于13年跟踪研究的判断与建议[J]. 广西师范大学学报(哲学社会科学版), 2017, 53(6): 108-114.

REN Y C. Current problems and solutions of rural left-behind children's education - A study based on 13 years of tracking observation[J]. Journal of Guangxi normal university (philosophy and social sciences edition), 2017, 53(6): 108-114.

[29] 罗国芬. 儿童权利视角: 农村留守儿童 "再问题化"[J]. 探索与争鸣, 2018(1): 79-83, 143.

LUO G F. A perspective of children's rights: Re-problemization of left-behind children in rural areas[J]. Exploration and free views, 2018(1): 79-83, 143.

[30] 叶敬忠, 王伊欢. 留守儿童的监护现状与特点[J]. 人口学刊, 2006, 28(3): 55-59.

YE J Z, WANG Y H. The Current situation and characteristics of care taking for children left behind[J]. Population journal, 2006, 28(3): 55-59.

[31] 范兴华, 方晓义. 不同监护类型留守儿童与一般儿童问题行为比较[J]. 中国临床心理学杂志, 2010, 18(2): 232-234, 237.

FAN X H, FANG X Y. A comparison of problem behaviors between parent-present children and parent-absent children in different care-taking modes[J]. Chinese journal of clinical psychology, 2010, 18(2): 232-234, 237.

[32] 刘晓慧, 王晓娟, 杨玉岩, 等. 不同监护类型留守儿童与一般儿童心理健康状况的比较研究[J]. 中国全科医学, 2012, 15(13): 1507-1510.

LIU X H, WANG X J, YANG Y Y, et al. Comparison of mental health status between left-behind children under different guardianships and non-left-behind children[J]. Chinese general practice, 2012, 15(13): 1507-1510.

[33] 蒋平. 农村留守儿童家庭教育基本缺失的问题及对策[J]. 理论观察, 2005(4): 79-81.

JIANG P. Problems and countermeasures of the basic lack of family education for left-behind children in rural areas[J]. Theory observe, 2005(4): 79-81.

[34] 冯璇坤. 理想·遮蔽·澄明:数字时代下留守儿童网络游戏现象的三重审视[J]. 理论月刊, 2020(12): 91-100.

FENG X K. Ideal, covering, and uncovering: A triple examination of the online game phenomenon of left-behind children in the digital age[J]. Theory monthly, 2020(12): 91-100.

[35] 杨国才, 朱金磊. 国内外留守儿童问题研究述评与展望[J]. 云南师范大学学报(哲学社会科学版), 2013, 45(5): 113-119.

YANG G C, ZHU J L. A summary review of the studies on the left-behind children both in China and in the world[J]. Journal of Yunnan normal university (humanities and social sciences), 2013, 45(5): 113-119.

[36] 陆继霞, 叶敬忠. 我国农村地区同辈群体对留守儿童的影响研究[J]. 农村经济, 2009(12): 104-107.

LU J X, YE J Z. Research on the influence of peer groups on left-behind children in rural areas of China[J]. Rural economy, 2009(12): 104-107.

[37] 杜尚荣, 刘芳. 乡村教师引领留守儿童重构精神生活的理论逻辑、责任担当与实现路径[J]. 教育科学, 2022, 38(3): 46-52.
DU S R, LIU F. The theoretical logic, responsibility and realization path of rural teachers leading left-behind children to reconstruct their spiritual life[J]. Education science, 2022, 38(3): 46-52.

[38] 赵瑞锁, 贾月. 校园真人秀节目对儿童社会化的积极影响[J]. 编辑之友, 2016(5): 56-59.
ZHAO R S, JIA Y. The positive effects of the campus reality show on children's socialization[J]. Editorial friend, 2016(5): 56-59.

[39] 施焦迪, 李红岩. 2007—2019 年我国留守儿童媒介素养研究现状探析[J]. 国际公关, 2021(1): 74-77.
SHI J D, LI H Y. Research status of media literacy of left-behind children in China from 2007 to 2019[J]. PR magazine, 2021(1): 74-77.

[40] 邓园. 苏北地区农村留守儿童媒介素养探究[J]. 传媒观察, 2018(1): 24-26.
DENG Y. Research on media literacy of rural left-behind children in northern Jiangsu Province[J]. Media observer, 2018(1): 24-26.

[41] 赵可云, 崔晓鸾, 杨鑫, 等. 大众媒介对农村留守儿童学习社会化影响的实证研究[J]. 现代远距离教育, 2018(3): 66-74.
ZHAO K Y, CUI X L, YANG X, et al. An empirical study on the influence of mass media on the socialization of left-behind children in rural areas[J]. Modern distance education, 2018(3): 66-74.

[42] 王清华, 郑欣. 数字代偿：智能手机与留守儿童的情感社会化研究[J]. 新闻界, 2022(3): 37-47, 94.
WANG Q H, ZHENG X. Digital compensation: A study on smart-phones and the emotional socialization of left-behind children[J]. Journalism and mass communication, 2022(3): 37-47, 94.

[43] MIRIC M, JIA N, HUANG K G. Using supervised machine learning for large-scale classification in management research: The case for identifying artificial intelligence patents[J]. Strategic management journal, 2023, 44(2): 491-519.

Twenty Years of Left-Behind Children Education in Rural China: Based on Structural Topic Model

WANG Xing[1], LI Yeye[2*], ZHOU Tianyu[3], LIU Feng[4]

（1. Chengdu Library and Information Center, Chinese Academy of Sciences, Chengdu 610299; 2. School of Management, Hainan University, Haikou 570100; 3. School of Artificial Intelligence, Hefei University of Technology, Xuancheng, China; 4. Suixi County Hancun Central School Xiaohu Primary School, Huaibei 235100）

Abstract: [Purpose/Significance] The introduction of national poverty alleviation policies and rural revitalization strategies has thrust the issue of education for left-behind children into the spotlight of scholarly attention. Education, far beyond serving as a mere instrument for personal growth and human capital accumulation for left-behind children, emerges as a pivotal measure in consolidating rural poverty alleviation endeavors and breaking the transmission of intergenerational poverty in China. It stands as a vital force propelling the future of rural revitalization. Yet, the existing literature on the education of left-behind children remains sporadic and dispersed. A more profound organizational effort, integrating, synthesizing, and evaluating this scattered literature, is imperative to establish a foundational framework for future research, fostering more cohesive and focused research endeavors. Presently, literature review studies primarily fall

DOI：10.13998/j.cnki.issn1002-1248.23-0691

into three categories: qualitative review methods, meta-analysis, and bibliometric analysis methods employing tools like Citespace. This study sets out to achieve a systematic and comprehensive understanding of education-related issues for rural left-behind children through text mining methods grounded in topic models. [Method/Process] The advent of artificial intelligence and machine learning technologies has empowered the processing and analysis of vast amounts of textual data. Previous research, employing latent dirichlet allocation (LDA) topic models, successfully mined texts related to teacher team construction reform policies, internationalization in higher education literature, news reports, and online comments. In this study, a corpus was meticulously constructed using abstract texts extracted from 2037 journal articles published between 2002 and 2023. The structural topic model (STM) was chosen for topic modeling, overcoming the limitations associated with LDA, with a specific emphasis on exploring the diversity and dynamism of topics within the existing literature. [Results/Conclusions] The culmination of this research effort identified eight distinct research themes: psychological well-being, factors leading to left-behind children, macro-level coping strategies, types of guardianship, review studies, family education, media literacy, and micro-level coping strategies. By synergizing document metadata information, the study systematically unraveled the evolving trends of these topics over time, providing crucial insights into potential shifts in the focus of left-behind children's education research. It is essential to note that this study, while collecting abstracts instead of full texts, may not capture the entirety of information contained in complete research articles. Future research endeavors should explore left-behind children's education more comprehensively, leveraging full-text mining techniques for a more nuanced understanding of this critical subject.

Keywords: left-behind children; topic model; structural topic model; information literacy